

ANALISIS PREDIKSI ANGKA PARTISIPASI SEKOLAH USIA 16–18 TAHUN DI JAWA BARAT MENGGUNAKAN RANDOM FOREST DAN FITUR RASIO PENDIDIKAN

Ilham Mukjizat, Asril Adi Sunarto, Agung Pambudi

Teknik Informatika, Universitas Muhammadiyah Sukabumi

Jl. R. Syamsudin, SH. No. 50 Kota Sukabumi Jawa Barat Indonesia 43113, Indonesia

ilham028@ummi.ac.id

ABSTRAK

Angka Partisipasi Sekolah (APS) usia 16–18 tahun merupakan indikator penting untuk melihat keberlanjutan partisipasi pendidikan pada fase pendidikan menengah. Provinsi Jawa Barat memiliki karakteristik wilayah yang beragam, sehingga capaian APS antar kabupaten/kota masih menunjukkan perbedaan. Penelitian ini bertujuan membangun model prediksi APS usia 16–18 tahun pada tingkat kabupaten/kota di Jawa Barat agar pola partisipasi sekolah dapat dianalisis secara lebih sistematis. Penelitian ini menggunakan pendekatan kuantitatif prediktif dengan kerangka CRISP-DM. Data yang digunakan berupa data panel 27 kabupaten/kota selama periode 2018–2025 dengan total 216 observasi. Variabel prediktor meliputi sosial-ekonomi, infrastruktur pendidikan, fitur rasio, fitur temporal, dan fitur spasial. Model yang digunakan adalah *Random Forest Regressor*, *Gradient Boosting Regressor*, dan *Support Vector Regression*, dengan evaluasi menggunakan MAE, RMSE, R^2 , serta validasi GroupKFold berbasis kabupaten/kota. Hasil penelitian menunjukkan bahwa model terbaik diperoleh dari *Random Forest Regressor* pada skenario Non-Spasial Rasio dengan MAE 4,3297, RMSE 5,4091, dan R^2 0,4522. Interpretasi *feature importance* dan SHAP menunjukkan bahwa IPM, rata-rata lama sekolah, rasio guru per 1.000 murid, dan rasio murid per guru menjadi fitur yang paling berkontribusi terhadap prediksi APS.

Kata kunci : Angka Partisipasi Sekolah, Random Forest, fitur rasio, SHAP, GroupKFold, Jawa Barat

1. PENDAHULUAN

Pendidikan merupakan salah satu faktor penting dalam pembangunan sumber daya manusia dan pengurangan ketimpangan antardaerah. Salah satu indikator yang dapat digunakan untuk melihat keberlanjutan partisipasi pendidikan adalah Angka Partisipasi Sekolah (APS). APS usia 16–18 tahun menggambarkan persentase penduduk pada kelompok umur tersebut yang masih bersekolah, sehingga indikator ini penting untuk melihat keberlanjutan pendidikan pada usia remaja atau fase pendidikan menengah. Partisipasi sekolah pada kelompok usia tersebut dapat dipengaruhi oleh berbagai faktor, seperti kemiskinan, kondisi sosial-ekonomi, akses layanan pendidikan, serta dukungan pembiayaan pendidikan [1], [2].

Provinsi Jawa Barat memiliki jumlah penduduk yang besar dan karakteristik wilayah yang beragam, baik dari sisi sosial-ekonomi, infrastruktur pendidikan, maupun kondisi geografis. Meskipun capaian APS usia 16–18 tahun menunjukkan kecenderungan meningkat, pemerataan antarwilayah masih menjadi persoalan. Beberapa penelitian sebelumnya menunjukkan bahwa partisipasi pendidikan menengah dapat dipengaruhi oleh indikator kapasitas layanan pendidikan, kondisi ekonomi wilayah, serta adanya pola dependensi spasial antarwilayah [3], [4].

Penelitian terkait angka putus sekolah dan partisipasi pendidikan juga menunjukkan bahwa faktor spasial dapat berperan dalam menjelaskan variasi indikator pendidikan antarwilayah [4], [5].

Permasalahan utama dalam penelitian ini adalah belum meratanya capaian APS usia 16–18 tahun antar kabupaten/kota di Provinsi Jawa Barat. Analisis

deskriptif hanya dapat menunjukkan wilayah dengan capaian APS tinggi atau rendah, tetapi belum cukup untuk membangun model prediksi yang mampu membaca pola berdasarkan kombinasi variabel sosial-ekonomi, infrastruktur pendidikan, rasio pendidikan, temporal, dan spasial. Oleh karena itu, diperlukan pendekatan prediktif yang dapat mempelajari pola dari data historis dan menghasilkan estimasi APS secara lebih sistematis.

Perkembangan metode *machine learning* memungkinkan data sosial-ekonomi dan pendidikan dianalisis secara lebih fleksibel, terutama ketika hubungan antarvariabel bersifat kompleks dan nonlinier [6].

Beberapa algoritma seperti Random Forest Regressor, Gradient Boosting Regressor, dan Support Vector Regression banyak digunakan dalam pemodelan prediktif karena mampu mengolah data tabular serta menangani pola hubungan antarfitur yang kompleks [7], [8], [9].

Dalam penelitian ini, pendekatan tersebut digunakan untuk memprediksi APS usia 16–18 tahun pada tingkat kabupaten/kota di Provinsi Jawa Barat.

Selain variabel sosial-ekonomi dan infrastruktur pendidikan, aspek spasial juga perlu diperhatikan dalam analisis kewilayahan. Kondisi suatu kabupaten/kota tidak selalu berdiri sendiri, karena wilayah yang berdekatan dapat memiliki keterkaitan sosial, ekonomi, maupun akses layanan pendidikan. Pendekatan spasial seperti *Queen Contiguity* dapat digunakan untuk membangun relasi ketetanggaan berdasarkan batas administratif wilayah [10].

Integrasi fitur spasial dalam pemodelan juga relevan karena pendekatan berbasis wilayah dapat

membantu menangkap pola keterkaitan geografis antarunit analisis [11], [12].

Namun, penggunaan jumlah absolut sekolah, guru, dan murid memiliki keterbatasan karena dapat dipengaruhi oleh ukuran wilayah dan jumlah penduduk atau murid. Wilayah dengan jumlah penduduk besar cenderung memiliki jumlah sekolah dan guru yang lebih banyak, tetapi belum tentu memiliki kapasitas layanan pendidikan yang lebih proporsional. Oleh karena itu, indikator rasio seperti rasio murid-guru dapat digunakan untuk menggambarkan kapasitas layanan pendidikan secara lebih relatif dan proporsional [13].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun model prediksi APS usia 16–18 tahun pada tingkat kabupaten/kota di Provinsi Jawa Barat dengan mengintegrasikan variabel sosial-ekonomi, infrastruktur pendidikan, fitur rasio, dimensi temporal, dan fitur spasial. Rencana penyelesaian masalah dilakukan melalui tahapan CRISP-DM, mulai dari pemahaman masalah, pemahaman data, prapemrosesan data, rekayasa fitur, pemodelan menggunakan beberapa algoritma *machine learning*, evaluasi model menggunakan MAE, RMSE, dan R^2 , serta interpretasi model menggunakan *feature importance* dan SHAP. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan model prediksi pendidikan berbasis *machine learning* serta menjadi alat bantu analitik awal dalam perencanaan pendidikan daerah.

2. TINJAUAN PUSTAKA

2.1. Angka Partisipasi Sekolah

Angka Partisipasi Sekolah (APS) merupakan indikator pendidikan yang menggambarkan persentase penduduk pada kelompok umur tertentu yang masih bersekolah. APS usia 16–18 tahun relevan untuk melihat keberlanjutan partisipasi pendidikan pada fase pendidikan menengah. Partisipasi sekolah dipengaruhi oleh berbagai faktor sosial-ekonomi, seperti kemiskinan, kondisi keluarga, dukungan pembiayaan pendidikan, serta ketersediaan layanan pendidikan [1], [2].

Penelitian pada pendidikan menengah juga menunjukkan bahwa jumlah sekolah, alokasi pendidikan, dan kondisi ekonomi wilayah memiliki hubungan dengan partisipasi pendidikan [3].

2.2. Machine Learning untuk Prediksi Pendidikan

Machine learning merupakan pendekatan komputasional yang digunakan untuk mempelajari pola dari data historis dan menghasilkan prediksi terhadap data baru. Dalam konteks pendidikan dan sosial-ekonomi, *machine learning* relevan digunakan karena mampu menangkap hubungan nonlinier dan interaksi kompleks antarvariabel [6], [7].

Penelitian ini menggunakan pendekatan *supervised learning* untuk tugas regresi karena variabel target yang diprediksi berupa nilai numerik, yaitu APS usia 16–18 tahun [6].

2.3. Ensemble Learning

Ensemble learning merupakan metode yang menggabungkan beberapa model dasar untuk meningkatkan stabilitas dan akurasi prediksi. *Random Forest* bekerja dengan membangun banyak pohon keputusan dari sampel data yang berbeda, kemudian menggabungkan hasil prediksi untuk menghasilkan prediksi akhir yang lebih stabil [7].

Sementara itu, *Gradient Boosting* membangun model secara bertahap dengan memperbaiki kesalahan model sebelumnya, sehingga mampu menangkap pola kompleks pada data tabular [8]. Sebagai pembanding, penelitian ini juga menggunakan *Support Vector Regression* (SVR), yang sesuai untuk pemodelan regresi nonlinier dengan bantuan fungsi kernel [9].

2.4. Fitur Rasio dan Fitur Spasial

Fitur rasio digunakan untuk menggambarkan kapasitas layanan pendidikan secara lebih proporsional. Jumlah absolut sekolah, guru, dan murid belum tentu cukup untuk membandingkan kondisi antarwilayah karena dipengaruhi oleh ukuran wilayah dan jumlah penduduk. Oleh karena itu, rasio murid-guru dan indikator sejenis dapat digunakan untuk melihat beban layanan pendidikan secara relatif [13].

Selain fitur rasio, fitur spasial juga digunakan karena kondisi suatu wilayah dapat berkaitan dengan wilayah sekitarnya. Pendekatan *Queen Contiguity* dapat digunakan untuk membentuk hubungan ketetanggaan berdasarkan batas wilayah yang saling bersinggungan [10].

Dalam kajian pendidikan, beberapa penelitian menunjukkan adanya dependensi spasial pada indikator partisipasi pendidikan dan angka putus sekolah [4], [5].

2.5. Evaluasi Model dan Interpretabilitas

Evaluasi model regresi dilakukan menggunakan metrik *Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). Pada data berbasis wilayah, skema validasi perlu memperhatikan struktur spasial atau pengelompokan wilayah agar hasil evaluasi tidak terlalu optimistis akibat kedekatan data latih dan data uji [14].

Selain evaluasi performa, interpretabilitas model juga penting agar hasil prediksi dapat dijelaskan. SHAP (*SHapley Additive exPlanations*) dapat digunakan untuk menjelaskan kontribusi setiap fitur terhadap hasil prediksi, sehingga model *machine learning* yang kompleks menjadi lebih transparan [15].

2.6. Penelitian Terdahulu

Beberapa penelitian terdahulu telah membahas partisipasi pendidikan, pendekatan spasial, dan *machine learning* pada data sosial-ekonomi. Pratama dan Amalia [3] meneliti determinan angka partisipasi kasar pendidikan menengah di Indonesia dan menunjukkan bahwa indikator pendidikan, ekonomi, serta kapasitas layanan pendidikan memiliki hubungan dengan partisipasi pendidikan.

Setianti et al. [4] menggunakan regresi spasial data panel untuk memodelkan angka partisipasi pendidikan menengah atas di Jawa Barat, sehingga menunjukkan bahwa pendekatan spasial relevan digunakan dalam analisis indikator pendidikan antarwilayah.

Nasir et al. [5] menerapkan *Spatial Autoregressive* (SAR) pada angka putus sekolah bagi anak usia wajib belajar di Provinsi Sulawesi Selatan dan menunjukkan adanya keterkaitan spasial dalam permasalahan pendidikan.

Selain itu, penelitian terkait integrasi *machine learning* dan data spasial menunjukkan bahwa informasi kewilayahan dapat digunakan untuk memperkuat analisis berbasis wilayah.

Firmansyah et al. [11] menerapkan pendekatan *machine learning-enhanced geographically weighted regression* untuk evaluasi spasial Indeks Pembangunan Manusia di wilayah Indonesia bagian barat.

Hartanto et al. [12] juga menunjukkan bahwa integrasi data geospasial dan *machine learning* dapat digunakan untuk pemodelan prediktif indikator sosial-ekonomi. Berdasarkan penelitian terdahulu tersebut, penelitian ini berbeda karena berfokus pada prediksi APS usia 16–18 tahun di Jawa Barat dengan membandingkan fitur dasar, fitur rasio, fitur spasial, serta kombinasi fitur rasio dan spasial menggunakan *Random Forest*, *Gradient Boosting*, dan *SVR*.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif prediktif dengan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Pendekatan ini digunakan karena tujuan penelitian adalah membangun model prediksi Angka Partisipasi Sekolah (APS) usia 16–18 tahun pada tingkat kabupaten/kota di Provinsi Jawa Barat. Tahapan penelitian meliputi pemahaman masalah, pengumpulan dan pemahaman data, prapemrosesan data, rekayasa fitur, pemodelan, evaluasi model, serta interpretasi hasil prediksi.

3.1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data panel 27 kabupaten/kota di Provinsi Jawa Barat selama periode 2018–2025, sehingga total observasi yang digunakan berjumlah 216 observasi. Variabel target dalam penelitian ini adalah APS usia 16–18 tahun. Variabel prediktor terdiri dari variabel sosial-ekonomi, infrastruktur pendidikan, fitur rasio, fitur temporal, dan fitur spasial.

Data APS dan variabel sosial-ekonomi diperoleh dari BPS Provinsi Jawa Barat, data infrastruktur pendidikan dilengkapi menggunakan Open Data Jawa Barat, sedangkan data geospasial batas administrasi kabupaten/kota diperoleh dari Batas Administrasi Indonesia GeoIT.

Variabel sosial-ekonomi yang digunakan meliputi persentase penduduk miskin, rata-rata lama

sekolah, Indeks Pembangunan Manusia (IPM), dan pengeluaran riil per kapita. Variabel infrastruktur pendidikan meliputi jumlah sekolah SMA/SMK, jumlah guru SMA/SMK, dan jumlah murid SMA/SMK. Selain itu, digunakan pula variabel tahun sebagai fitur temporal untuk menangkap pola perubahan antarwaktu.

3.2. Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk memastikan data siap digunakan dalam proses pemodelan. Proses ini meliputi penggabungan data berdasarkan kode wilayah dan tahun, penyesuaian nama kabupaten/kota, pemeriksaan *missing value*, pemeriksaan duplikasi data, serta penyelarasan data tabular dengan data geospasial. Data infrastruktur pendidikan yang memiliki kekosongan pada sumber utama dilengkapi menggunakan sumber resmi pelengkap agar struktur data panel tetap lengkap.

Data geospasial batas administrasi kabupaten/kota digunakan untuk membentuk relasi ketetanggaan antarwilayah. Data tabular dan data geospasial disesuaikan menggunakan kode wilayah agar setiap kabupaten/kota dalam dataset memiliki pasangan geometri yang sesuai. Setelah proses validasi, dataset akhir digunakan untuk tahap rekayasa fitur dan pemodelan.

3.3. Rekayasa Fitur

Rekayasa fitur dilakukan dengan membentuk fitur rasio dan fitur spasial. Fitur rasio dibentuk dari variabel jumlah sekolah, jumlah guru, dan jumlah murid untuk menggambarkan kapasitas layanan pendidikan secara lebih proporsional antarwilayah. Fitur rasio yang digunakan meliputi rasio murid per sekolah, rasio guru per sekolah, rasio murid per guru, rasio guru per 1.000 murid, dan rasio sekolah per 10.000 murid.

Fitur spasial dibentuk menggunakan pendekatan *Queen Contiguity*. Dalam pendekatan ini, dua wilayah dianggap bertetangga apabila saling bersinggungan melalui sisi atau titik sudut. Berdasarkan matriks ketetanggaan tersebut, dibentuk beberapa fitur spasial, yaitu jumlah tetangga, status perbatasan, *spatial lag* pada variabel prediktor, dan kluster spasial LISA (*Local Indicators of Spatial Association*). Penelitian ini tidak menggunakan *spatial lag* APS pada tahun yang sama untuk menghindari risiko *data leakage*, karena APS merupakan variabel target.

3.4. Skenario Pemodelan

Pemodelan dilakukan dengan membandingkan empat skenario fitur. Skenario A atau Non-Spasial Dasar menggunakan variabel sosial-ekonomi, variabel infrastruktur pendidikan absolut, dan tahun. Skenario B atau Non-Spasial Rasio menambahkan fitur rasio infrastruktur pendidikan ke dalam fitur dasar. Skenario C atau Spasial Dasar menambahkan fitur spasial dasar, seperti *spatial lag* kemiskinan, LISA cluster, jumlah tetangga, dan status perbatasan. Skenario D atau

Spasial Lengkap menggabungkan fitur rasio, fitur spasial dasar, dan *spatial lag* pada beberapa variabel prediktor.

Model yang digunakan dalam penelitian ini adalah *Random Forest Regressor*, *Gradient Boosting Regressor*, dan *Support Vector Regression* (SVR) dengan kernel RBF. Ketiga model tersebut dipilih karena dapat digunakan untuk pemodelan regresi prediktif dan mampu menangani data tabular dengan pola hubungan antarvariabel yang kompleks.

3.5. Evaluasi Model

Evaluasi model dilakukan menggunakan skema GroupKFold dengan grup berdasarkan identitas kabupaten/kota. Skema ini digunakan agar seluruh observasi tahunan dari satu kabupaten/kota berada pada kelompok data yang sama, sehingga pengujian model lebih mencerminkan kemampuan generalisasi antarwilayah. Pendekatan ini juga digunakan untuk mengurangi risiko bias evaluasi akibat observasi berulang pada wilayah yang sama.

Metrik evaluasi yang digunakan adalah *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). MAE digunakan untuk mengukur rata-rata kesalahan absolut prediksi, RMSE digunakan untuk mengukur kesalahan prediksi dengan penalti lebih besar terhadap galat yang tinggi, sedangkan R^2 digunakan untuk mengetahui proporsi variasi data yang dapat dijelaskan oleh model. Model terbaik dipilih berdasarkan nilai MAE dan RMSE yang paling rendah serta nilai R^2 yang lebih tinggi.

3.6. Interpretasi Model

Setelah model terbaik diperoleh, interpretasi dilakukan menggunakan *feature importance* dan SHAP (*SHapley Additive exPlanations*). *Feature importance* digunakan untuk mengetahui fitur yang memiliki kontribusi besar dalam proses prediksi model, sedangkan SHAP digunakan untuk menjelaskan kontribusi fitur terhadap arah dan besarnya perubahan prediksi. Interpretasi ini dilakukan agar model tidak hanya menghasilkan prediksi, tetapi juga dapat memberikan informasi mengenai faktor-faktor yang berperan penting dalam prediksi APS.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Prapemrosesan Data

Dataset yang digunakan dalam penelitian ini merupakan data panel 27 kabupaten/kota di Provinsi Jawa Barat selama periode 2018–2025. Dengan demikian, total observasi yang digunakan berjumlah 216 observasi. Variabel target dalam penelitian ini adalah Angka Partisipasi Sekolah (APS) usia 16–18 tahun, sedangkan variabel prediktor terdiri atas variabel sosial-ekonomi, infrastruktur pendidikan, fitur rasio, fitur temporal, dan fitur spasial. Hasil pemeriksaan data menunjukkan bahwa dataset akhir tidak memiliki *missing value* pada variabel yang

digunakan dalam pemodelan, sehingga data dapat digunakan untuk proses pelatihan dan evaluasi model.

Tabel 1. Ringkasan dataset penelitian

Komponen Dataset	Keterangan
Unit analisis	Kabupaten/kota di Provinsi Jawa Barat
Jumlah wilayah	27 kabupaten/kota
Periode data	2018–2025
Jumlah tahun pengamatan	8 tahun
Total observasi	216 observasi
Variabel target	APS usia 16–18 tahun
Jenis model	Regresi prediktif
Struktur data	Panel <i>long format</i>
<i>Missing value</i>	Tidak ditemukan

Berdasarkan Tabel 1, dataset penelitian memiliki struktur panel yang seimbang karena setiap kabupaten/kota memiliki observasi lengkap selama delapan tahun pengamatan. Struktur ini memungkinkan model mempelajari pola APS antarwilayah dan antarwaktu. Selain itu, data tabular juga telah diselaraskan dengan data geospasial administratif, sehingga dapat digunakan untuk membentuk fitur spasial berbasis ketetanggaan wilayah.

4.2. Hasil Rekayasa Fitur

Rekayasa fitur dilakukan untuk memperkaya representasi data sebelum proses pemodelan. Fitur rasio dibentuk dari variabel jumlah sekolah, jumlah guru, dan jumlah murid SMA/SMK. Fitur ini digunakan untuk menggambarkan kapasitas layanan pendidikan secara lebih proporsional antarwilayah. Fitur rasio yang dibentuk meliputi rasio murid per sekolah, rasio guru per sekolah, rasio murid per guru, rasio guru per 1.000 murid, dan rasio sekolah per 10.000 murid.

Selain fitur rasio, penelitian ini juga membentuk fitur spasial menggunakan pendekatan *Queen Contiguity*. Hasil pembentukan matriks ketetanggaan menunjukkan bahwa seluruh kabupaten/kota memiliki tetangga, sehingga tidak terdapat wilayah yang teridentifikasi sebagai *island*. Fitur spasial yang dihasilkan meliputi jumlah tetangga, status perbatasan, *spatial lag* variabel prediktor, dan klaster spasial LISA. Pembentukan fitur spasial dilakukan tanpa menggunakan *spatial lag* APS pada tahun yang sama untuk menghindari risiko *data leakage*.

4.3. Hasil Pemodelan Empat Skenario

Pemodelan dilakukan dengan membandingkan empat skenario fitur, yaitu Non-Spasial Dasar, Non-Spasial Rasio, Spasial Dasar, dan Spasial Lengkap. Setiap skenario diuji menggunakan tiga algoritma, yaitu *Random Forest Regressor*, *Gradient Boosting Regressor*, dan *Support Vector Regression*. Evaluasi dilakukan menggunakan metrik MAE, RMSE, dan R^2 . Hasil perbandingan performa model ditampilkan pada Tabel 2.

Tabel 2. Hasil evaluasi model pada empat skenario

Skenario	Model	MAE	RMSE	R ²
Non-Spasial Rasio	Random Forest	4,4209	5,4654	0,4449
Non-Spasial Rasio	Gradient Boosting	4,6796	5,7314	0,3936
Spasial Lengkap	Gradient Boosting	4,7828	5,8313	0,3736
Spasial Lengkap	Random Forest	4,8621	5,9080	0,3539
Non-Spasial Dasar	Random Forest	4,7886	5,9549	0,3261
Non-Spasial Dasar	Gradient Boosting	4,8724	5,9646	0,3306
Non-Spasial Rasio	SVR	4,7148	6,1283	0,3088
Spasial Dasar	Random Forest	4,9111	6,1565	0,2688
Spasial Dasar	Gradient Boosting	4,8715	6,1903	0,2785
Spasial Lengkap	SVR	4,8641	6,2108	0,2916
Non-Spasial Dasar	SVR	4,8983	6,3149	0,2427
Spasial Dasar	SVR	5,2757	6,8543	0,1317

Berdasarkan Tabel 2, skenario Non-Spasial Rasio dengan algoritma *Random Forest Regressor* menghasilkan performa terbaik dibandingkan skenario lainnya, dengan MAE 4,4209, RMSE 5,4654, dan R² 0,4449. Hasil ini menunjukkan bahwa penambahan fitur rasio infrastruktur pendidikan mampu meningkatkan performa prediksi dibandingkan penggunaan variabel dasar saja. Sementara itu, fitur spasial tetap memberikan nilai analitis, tetapi belum mampu melampaui performa skenario Non-Spasial Rasio pada dataset penelitian ini.

4.4. Hasil Hyperparameter Tuning

Setelah diperoleh hasil awal dari empat skenario, dilakukan *hyperparameter tuning* pada model dengan performa terbaik. Tuning dilakukan untuk mencari kombinasi parameter yang lebih optimal sehingga dapat meningkatkan performa prediksi. Hasil tuning model terbaik ditampilkan pada Tabel 3.

Tabel 3. Hasil hyperparameter tuning model terbaik

Skenario	Model	MAE	RMSE	R ²
Non-Spasial Rasio	Random Forest	4,3297	5,4091	0,4522
Non-Spasial Rasio	Gradient Boosting	4,4599	5,4515	0,4510
Spasial Lengkap	Gradient Boosting	4,4708	5,5749	0,4233
Spasial Lengkap	Random Forest	4,6232	5,6878	0,4015

Berdasarkan Tabel 3, model terbaik setelah proses tuning adalah *Random Forest Regressor* pada skenario Non-Spasial Rasio, dengan MAE sebesar 4,3297, RMSE sebesar 5,4091, dan R² sebesar 0,4522. Nilai MAE tersebut menunjukkan bahwa rata-rata kesalahan prediksi model berada pada kisaran 4,33 poin APS. Nilai R² sebesar 0,4522 menunjukkan bahwa model mampu menjelaskan sekitar 45,22% variasi APS pada skema validasi yang digunakan. Hasil ini cukup relevan mengingat data sosial-ekonomi wilayah memiliki karakteristik kompleks dan jumlah observasi yang terbatas.

4.5. Interpretasi Model

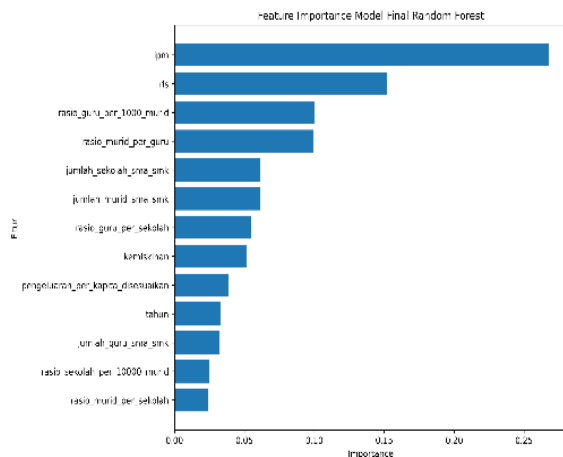
Interpretasi model dilakukan menggunakan *feature importance* dan SHAP untuk mengetahui fitur yang paling berkontribusi terhadap hasil prediksi APS. Hasil *feature importance* menunjukkan bahwa Indeks Pembangunan Manusia (IPM), rata-rata lama sekolah (RLS), rasio guru per 1.000 murid, dan rasio murid per guru merupakan fitur dengan kontribusi terbesar dalam model.

Tabel 4. Fitur dengan kontribusi terbesar pada model final

Peringkat	Fitur	Feature Importance
1	IPM	0,2676
2	Rata-rata lama sekolah	0,1521
3	Rasio guru per 1.000 murid	0,1002
4	Rasio murid per guru	0,0991
5	Jumlah sekolah SMA/SMK	0,0612
6	Jumlah murid SMA/SMK	0,0609
7	Rasio guru per sekolah	0,0547
8	Persentase penduduk miskin	0,0515

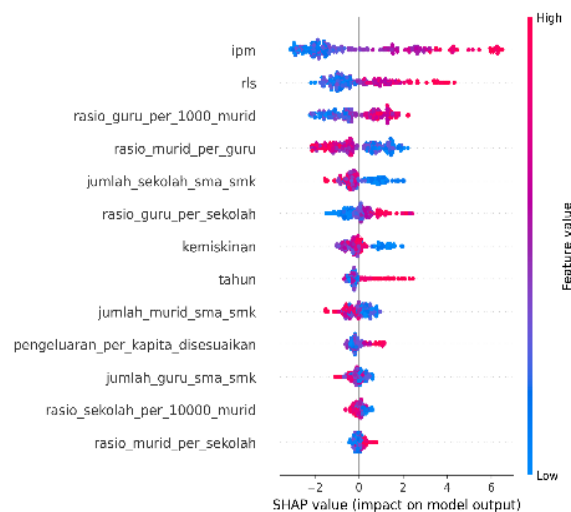
Berdasarkan Tabel 4, IPM menjadi fitur yang paling dominan dalam prediksi APS. Hal ini menunjukkan bahwa kondisi pembangunan manusia secara umum memiliki informasi prediktif yang kuat terhadap partisipasi sekolah usia 16–18 tahun. Selain itu, RLS juga menjadi fitur penting karena mencerminkan capaian pendidikan masyarakat. Fitur rasio pendidikan, terutama rasio guru per 1.000 murid dan rasio murid per guru, menunjukkan bahwa kapasitas layanan pendidikan secara relatif memiliki peran penting dalam prediksi APS.

Hasil SHAP memperkuat temuan tersebut. IPM, RLS, rasio guru per 1.000 murid, dan rasio murid per guru menjadi fitur dengan nilai kontribusi global terbesar. Secara umum, nilai IPM dan RLS yang lebih tinggi cenderung mendorong prediksi APS menjadi lebih tinggi. Sebaliknya, rasio murid per guru yang tinggi cenderung berkaitan dengan prediksi APS yang lebih rendah, karena rasio tersebut menggambarkan beban layanan guru yang lebih besar. Namun, interpretasi ini bersifat prediktif dan tidak dapat langsung dimaknai sebagai hubungan sebab-akibat.



Gambar 1. Feature importance model final

Gambar 1 menunjukkan urutan fitur berdasarkan kontribusinya dalam model *Random Forest Regressor*. Dari gambar tersebut dapat dilihat bahwa IPM, RLS, dan fitur rasio pendidikan menjadi kelompok fitur yang dominan dalam proses prediksi APS.



Gambar 2. SHAP summary plot model final

Gambar 2 menunjukkan arah kontribusi setiap fitur terhadap hasil prediksi. Titik pada sisi kanan menunjukkan kontribusi fitur dalam meningkatkan prediksi APS, sedangkan titik pada sisi kiri menunjukkan kontribusi fitur dalam menurunkan prediksi APS.

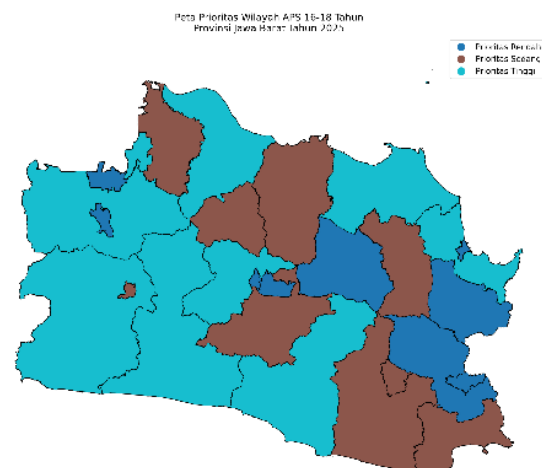
4.6. Peta Prioritas Wilayah APS Tahun 2025

Selain menghasilkan model prediksi, penelitian ini juga menyusun peta prioritas wilayah APS tahun 2025. Peta prioritas disusun berdasarkan kombinasi APS aktual dan APS prediksi, sehingga dapat memberikan gambaran wilayah yang relatif membutuhkan perhatian lebih besar. Kategori prioritas dibagi menjadi tiga, yaitu Prioritas Tinggi, Prioritas Sedang, dan Prioritas Rendah.

Tabel 5. Wilayah kategori prioritas tinggi tahun 2025

No	Kabupaten/Kota	APS Aktual	APS Prediksi
1	Kabupaten Sukabumi	68,11	66,43
2	Kabupaten Bogor	68,38	70,18
3	Kabupaten Bandung Barat	68,64	74,75
4	Kabupaten Cianjur	72,53	69,01
5	Kabupaten Karawang	71,92	73,25
6	Kabupaten Indramayu	70,33	75,80
7	Kabupaten Garut	71,26	74,60
8	Kabupaten Cirebon	74,09	73,62
9	Kota Bekasi	72,08	79,82

Berdasarkan Tabel 5, beberapa wilayah yang termasuk kategori Prioritas Tinggi antara lain Kabupaten Sukabumi, Kabupaten Bogor, Kabupaten Bandung Barat, Kabupaten Cianjur, Kabupaten Karawang, Kabupaten Indramayu, Kabupaten Garut, Kabupaten Cirebon, dan Kota Bekasi. Wilayah tersebut memiliki nilai APS aktual dan/atau prediksi yang relatif lebih rendah dibandingkan wilayah lain, sehingga dapat menjadi perhatian awal dalam perencanaan pendidikan daerah. Namun, kategori prioritas ini bersifat relatif antarwilayah dan tidak dimaksudkan sebagai keputusan kebijakan final.



Gambar 3. Peta prioritas wilayah APS tahun 2025

Gambar 3 menunjukkan sebaran spasial kategori prioritas APS tahun 2025. Peta ini berfungsi sebagai alat bantu visual untuk melihat konsentrasi wilayah yang memerlukan perhatian lebih lanjut dalam peningkatan partisipasi sekolah usia 16–18 tahun.

4.7. Pembahasan

Hasil penelitian menunjukkan bahwa fitur rasio infrastruktur pendidikan memberikan kontribusi penting terhadap peningkatan performa prediksi. Hal ini terlihat dari skenario Non-Spasial Rasio yang menghasilkan performa terbaik dibandingkan skenario lainnya. Temuan ini menunjukkan bahwa perbandingan kapasitas layanan pendidikan secara relatif lebih informatif dibandingkan hanya

menggunakan jumlah absolut sekolah, guru, dan murid.

Integrasi fitur spasial dalam penelitian ini tetap memiliki nilai penting karena mampu merepresentasikan konteks ketetanggaan dan hubungan antarwilayah. Namun, hasil evaluasi menunjukkan bahwa fitur spasial belum memberikan peningkatan performa yang lebih baik dibandingkan fitur rasio. Dengan demikian, hasil penelitian ini tidak menunjukkan bahwa fitur spasial tidak relevan, melainkan menunjukkan bahwa pada dataset dan skenario yang digunakan, fitur rasio memiliki informasi prediktif yang lebih kuat.

Secara keseluruhan, model *Random Forest Regressor* pada skenario Non-Spasial Rasio dapat digunakan sebagai model prediksi terbaik dalam penelitian ini. Hasil interpretasi juga menunjukkan bahwa APS usia 16–18 tahun berkaitan secara prediktif dengan capaian pembangunan manusia, capaian pendidikan masyarakat, dan kapasitas layanan pendidikan. Oleh karena itu, hasil model dan peta prioritas dapat digunakan sebagai alat bantu analitik awal dalam mendukung perencanaan pendidikan berbasis data.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, model prediksi Angka Partisipasi Sekolah (APS) usia 16–18 tahun pada tingkat kabupaten/kota di Provinsi Jawa Barat berhasil dibangun menggunakan pendekatan *ensemble machine learning* berbasis data panel tahun 2018–2025. Hasil pengujian menunjukkan bahwa model terbaik diperoleh dari algoritma *Random Forest Regressor* pada skenario Non-Spasial Rasio dengan nilai MAE sebesar 4,3297, RMSE sebesar 5,4091, dan R^2 sebesar 0,4522. Hasil tersebut menunjukkan bahwa fitur rasio infrastruktur pendidikan memberikan kontribusi penting dalam meningkatkan performa prediksi dibandingkan penggunaan variabel infrastruktur dalam bentuk jumlah absolut. Interpretasi model menggunakan *feature importance* dan SHAP menunjukkan bahwa IPM, rata-rata lama sekolah, rasio guru per 1.000 murid, dan rasio murid per guru merupakan fitur yang paling berkontribusi terhadap prediksi APS. Integrasi fitur spasial tetap memiliki nilai analitis karena mampu merepresentasikan konteks ketetanggaan antarwilayah, meskipun pada penelitian ini belum mampu melampaui performa skenario berbasis fitur rasio. Penelitian ini juga menghasilkan peta prioritas wilayah APS tahun 2025 yang dapat digunakan sebagai alat bantu analitik awal dalam perencanaan pendidikan daerah. Untuk penelitian selanjutnya, disarankan agar cakupan data diperluas pada tingkat kecamatan, sekolah, atau rumah tangga, serta mempertimbangkan variabel aksesibilitas pendidikan seperti jarak ke sekolah, waktu tempuh, kondisi transportasi, biaya pendidikan, dan angka putus sekolah. Selain itu, pengembangan fitur spasial dapat diperluas dengan pendekatan berbasis jarak, jaringan jalan, atau waktu tempuh agar representasi

spasial yang digunakan lebih sesuai dengan konteks akses layanan pendidikan.

DAFTAR PUSTAKA

- [1] N. R. Al Zaryah, B. Robiani, S. Rohima, A. Yulianita, and Y. Anggraini, “Pengaruh Tingkat Kemiskinan dan Putus Sekolah terhadap Partisipasi Angkatan Kerja di Provinsi Sumatera Selatan,” *J. Pengabd. Masy. dan Ris. Pendidik.*, vol. 4, no. 2, pp. 10985–10995, 2025, doi: 10.31004/jerkin.v4i2.3430.
- [2] S. Rochman, F. Indriyani, S. D. Sapitri, P. Anggreini, and S. Yuliani, “Peran Pendidikan Berkualitas dalam Memutus Kemiskinan untuk Mendukung SDGs di Indonesia,” *J. Penelit. Ilmu Pendidik. Indones.*, vol. 4, no. 3, pp. 1503–1512, 2025, doi: 10.31004/jpion.v4i3.585.
- [3] M. S. I. Pratama and F. Amalia, “Determinan Angka Partisipasi Kasar Pendidikan Menengah di Indonesia,” *J. Ekono Insentif*, vol. 18, no. 2, pp. 94–106, 2024, doi: 10.36787/jei.v18i2.1757.
- [4] D. Setianti, R. Arisanti, and Y. Suparman, “Regresi Spasial Data Panel Untuk Pemodelan Angka Partisipasi Pendidikan Menengah Atas di Jawa Barat,” *BIASTATISTICS J. Stat. THEORY Apl.*, vol. 5, no. 1, pp. 1–12, 2024.
- [5] R. Nasir, S. Annas, and M. Nusrang, “Pemodelan dengan Spatial Autoregressive (SAR) pada Angka Putus Sekolah Bagi Anak Usia Wajib Belajar di Provinsi Sulawesi Selatan,” *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 3, no. 1, pp. 44–50, 2021, doi: 10.35580/variansiunm9358.
- [6] R. S. Nurhalizah, R. Ardianto, and Purwono, “Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review,” *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 61–72, 2024, doi: 10.54082/jiki.168.
- [7] D. N. Handayani and S. Qutub, “Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan,” *J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 406–412, 2025, doi: 10.31004/riggs.v4i2.512.
- [8] R. Dahlia, L. A. Fitriana, and S. Seimahuira, “ANALISIS ALGORITMA GRADIENT BOOSTING DALAM PENGARUH MASYARAKAT MEMILIH RUMAH SEWA,” *J. Manaj. Inform. Sist. Inf.*, vol. 8, no. 1, pp. 36–44, 2025.
- [9] N. P. S. Y. Artini, I. W. Sumarjaya, and D. P. E. Nilakusmawati, “PENERAPAN METODE SUPPORT VECTOR REGRESSION (SVR) DENGAN ALGORITMA GRID SEARCH DALAM PERAMALAN HARGA SAHAM Ni,” *E-Jurnal Mat.*, vol. 13, no. 2, pp. 94–104, 2024, doi: 10.24843/MTK.2024.v13.i02.p447.
- [10] G. Fajri, S. Syafriandi, N. Amalita, and Z. Martha, “Comparison of Queen Contiguity and Customized Weighting Matrices on Spatial Regression to Identify Factors Impacting Poverty in East Java,” *UNP J. Stat. DATA Sci.*, vol. 1, no.

- 3, pp. 203–210, 2024, doi: 10.24036/ujsds/vol1-iss3/67.
- [11] G. A. Firmansyah, J. Zeniarja, H. Al Azies, S. Winarno, and S. P. Ganiswari, “Machine Learning-Enhanced Geographically Weighted Regression for Spatial Evaluation of Human Development Index across Western Indonesia,” *JAGI J. Appl. GEOSPATIAL Inf.*, vol. 7, no. 2, 2023, doi: 10.30871/jagi.v7i2.6755.
- [12] M. B. Hartanto, H. D. Yunita, F. Fahurian, T. M. F. H.S, Desmon, and R. F. Purnomo, “Integration of Geo-Spatial Data and Machine Learning for Socio-Economic Forecasting,” *J. Appl. Informatics Comput.*, vol. 10, no. 2, pp. 2025–2034, 2026, doi: 10.30871/jaic.v10i2.12234.
- [13] F. S. Ahmad, “Perilaku Belanja Pegawai Fungsi Pendidikan dan Output Layanan : Analisis Rasio Murid-Guru SD dan SMP di Pulau Jawa,” *J. Ekon. dan Kebijak. Pembang.*, vol. 14, no. 2, pp. 133–150, 2025, doi: 10.29244/jekp.14.2.2025.133-150.
- [14] D. Radočaj, I. Plaščak, and M. Jurišič, “Comparative Assessment of Regular and Spatial Cross-Validation in Subfield Machine Learning Prediction of Maize Yield from Sentinel-2 Phenology,” *Eng*, vol. 6, no. 270, 2025, doi: 10.3390/eng6100270.
- [15] V. A. P.- Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, “Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development,” *Clin. Transl. Sci.*, vol. 17, no. 11, pp. 1–15, 2024, doi: 10.1111/cts.70056.